(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 May 2002 (02.05.2002)

PCT

(10) International Publication Number
WO 02/34909 A2

(51) International Patent Classification[7]: C12N 15/11, C07K 14/16, G01N 33/569

(21) International Application Number: PCT/US01/48040

(22) International Filing Date: 26 October 2001 (26.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/698,311 27 October 2000 (27.10.2000) US

(71) Applicant: ABBOTT LABORATORIES [US/US]; 100 Abbott Park Road, CHAD 0377/AP6D-2, Abbott Park, IL 60064-6050 (US).

(72) Inventors: STEWART, Kent, D.; 4715 Kings Way North, Gurnee, IL 60031 (US). STEFFY, Kevin, R.; 9050 Carmino Santa Fe, San Diego, CA 92121 (US). KEMPF, Dale, J.; 256 Tyler Court, Libertyville, IL 60048 (US). HARRIS, Kevin, S.; 920 N. Sterling Avenue, Apt. 214, Palatine, IL 60067 (US). HUTH, Jeffrey, R.; 1103 Tracy Lane, Libertyville, IL 60048 (US). STOLL, Vincent, S.; 218 E. Winchester Road, Libertyville, IL 60048 (US). HARLAN, John, E.; 689 Portree Lane, Lake Zurich, IL 60047 (US). NG, Iok, C.; 3229 N. Walker E., Arlington Heights, IL 60004 (US). BETZ, Stephen, F.; 1136 Lamb Lane, Gurnee, IL 60031 (US).

(74) Agents: CASUTO, Dianne et al.; Abbott Laboratories, 100 Abbott Park Road, CHAD 0377/AP6D-2, Abbott Park, IL 60064-6050 (US).

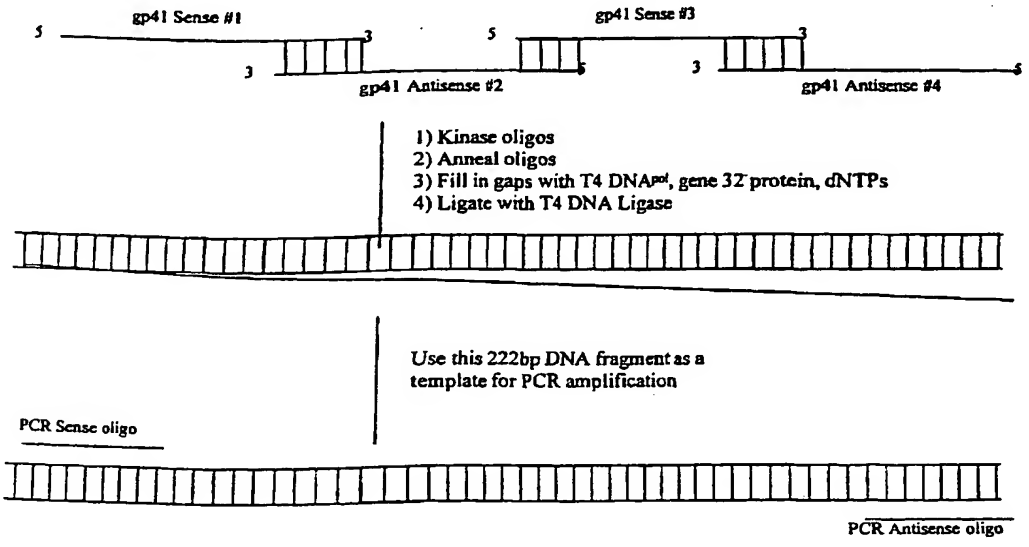
(81) Designated States *(national)*: CA, JP, MX.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: ENGINEERED CHIMERA OF PROTEIN FRAGMENTS AND METHODS OF USE THEREOF



(57) Abstract: The subject invention encompasses novel proteins related to the human immunodeficiency virus (HIV-1) gp41 protein and to methods of use thereof. For example, the proteins may be utilized in the screening of anti-HIV compounds.

WO 02/34909 A2

ENGINEERED CHIMERA OF PROTEIN FRAGMENTS AND METHODS OF USE THEREOF

BACKGROUND OF THE INVENTION

Technical Field

The subject invention encompasses novel proteins related to the human immunodeficiency virus (HIV-1) gp41 protein and to methods of use thereof. For example, the proteins may be utilized in the screening of anti-HIV compounds.

Background Information

The causative agent for Acquired Immune Deficiency Syndrome (A.I.D.S.) is an enveloped virus called human immunodeficiency virus type 1 (HIV-1). The viral envelope contains a protein complex that is vital for viral entry into susceptible cells. This envelope complex specifies which cells to infect and mediates the fusion of the viral membrane with the host plasma membrane allowing the invasion of the viral genome. It is also responsible for syncytium formation that occurs when an infected cell fuses with a neighboring uninfected cell.

The envelope complex is composed of two viral proteins, gp120 and gp41. These two proteins come from a common precursor termed gp160, which is cleaved by a cellular convertase to generate gp120 and gp41 (Fields, B., Virology 1996, Lippincott-Raven Publishers, Philadelphia, pp. 1881-1952). The cleaved products remain noncovalently associated with each other in an oligomeric form (gp120/gp41) on the surface of the virion. Binding of gp120/gp41 to a host receptor, CD4, in conjunction with either one of two chemokine co-receptors, termed CCR-5 and CXCR-4, causes a conformational change in the viral

gp120/gp41 complex where gp41 mediates the fusion of the viral membrane with the host membrane.

The gp41 protein is a transmembrane protein with a multifaceted ectodomain. At its N-terminus resides a hydrophobic fusion peptide, which is crucial for membrane fusion. Next are two 4,3 hydrophobic (heptad) repeats which are involved in the formation of coiled-coils. The N-terminal heptad repeat is termed N-helix and the C-terminal heptad repeat is called C-helix. A loop region is present between these two helices. The heptad regions form a helical trimer of antiparallel dimmers (Lu et al. 1995 Nat. Struct. Biol. 2, pp. 1075-1082).

The crystal structure of gp41 demonstrates that the gp41 core is a six-helix bundle formed by three molecules of gp41 in which the N and C helices are arranged into three hairpins (Chan et al. 1997 PNAS 94, 14036-14313; Weissenhorn et al. 1997 Nature 387, 426-430). The inner N-helices form the trimeric-coiled coil with the C-helices packing in an antiparallel manner into three, well-conserved hydrophobic grove along the outside of the coiled coil. In this state, gp41 is in the fusion-active conformation by bringing the viral and host membranes in close juxtaposition, overcoming the energy barrier for membrane fusion (Chan et al 1997 PNAS 94, 14036-14313; Furuta et al. 1998 Nat. Struct. Biol. 5, 276-279; Hughson 1997 Curr. Biol. 7, R565-R569; Weissnhorn et al. 1997 Nature 387, 426-430).

Prior to binding to the host cell, gp41 is said to be in a nonfusogenic state but, upon binding to CD4/DDR-5 or CD4/CXCR-4, a conformational change occurs in gp41 transforming it into the fusogenic state. The hydrophobic N-terminal peptide of gp41 inserts into the host membrane.

In this situation, gp41 resides in both the viral membrane and the host membrane. This transient molecular species is called the prehairpin intermediate (Munoz-Barroso et al. 1998 J. Cell Biol. 140, 315-323; Chan and Kim 1998 Cell 93, 681-684; Furuta et al. 1998 Nat. Struct. Biol. 5, 276-279; Jones et al. 1998 J. Biol. Chem. 273, 404-409). Association of the C-helix with the N-helix into a trimer of helices represents the fusion active state of gp41. Synthetic peptides representing the C-helix are found to be inhibitors of HIV infection and syncytia formation at nanomolar concentrations in cell culture experiments (Jiang et al. 1993 Nature 365, 113; Wild et al. 1994 P.N.A.S. 91, 12676-12680; Lu et al. Nat. Struct. Biol. 2, 1075-1082; Chan et al. 1998 P.N.A.S. 95, 15613-15617; Rimsky et al. 1998 J. Virol. 72, 986-993). These peptides bind to the transiently exposed N-helix coiled coil in the prehairpin intermediate. One of these peptides, T20, has shown to exhibit antiviral activity in humans (Kirby et al. 1998 Nat. Med. 4, 1302-1307) showing the utility of this treatment strategy.

The crystal structure of the N-terminal coiled coil trimer reveals three symmetrically located pockets on its surface which bind a conserved motif of Trp-Trp-Ile found in the C-peptide. Peptide mimics of the Trp-Trp-Ile motif (Eckert et al. 1999 Cell 99, 103-115; see also WO/06599) show anti-fusion and antiviral activity. It is therefore desirable to discover small organic molecules which can bind the Trp-Trp-Ile binding pocket and which may serve as leads in drug discovery. The proteins of the present invention provide for such discovery.

Computational methods to discover such compounds have been disclosed (Debnath et al. 1999 J. Med. Chem. 42, 3203-

3209). This computation is derived from studies from protein with occupied Trp-Trp-Ile pockets. However, proteins with unoccupied Trp-Trp-Ile pocket would be more useful for drug discovery efforts. One such protein containing only 17 residues of the HIV-1 gp41 N-helix fused to a 29-residue yeast GCN4 protein analog has been described (Eckert et al. 1999 Cell 99, 103-115). The proteins described in the present invention differ significantly from this yeast construct in that residues from both N- and C-peptide portions HIV-1 gp41 are used and present an unoccupied Trp-Trp-Ile pocket useful for drug discovery. Furthermore, the Eckert et al. protein has an exposed N-helical region of GCN4 that may engender unanticipated and/or undesirable binding properties (see also WO 00/55377). The N-helical portion of the proteins described herein, in contrast, have only their Trp-Trp-Ile pocket exposed, with the remainder of the N-helix protected by an attached C-helix region. Therefore, the proteins of the present invention will not have any unanticipated and/or undesirable binding properties that might be observed in connection with the Eckert et al. protein (see also WO 00/55377).

Additionally, there are several reasons why the pocket (i.e., Trp-Trp-Ile) is important for targeting drugs. First, mutagenesis studies show N-peptide residues forming the pocket are critical for membrane fusion (Dubay et al. 1992 J. Virol. 66, 4748-4756; Cao et al. 1993 J. Virol. 72, 2747-2755; Chen et al. 1993 J. Virol. 67, 3615-3619; Wild et al. 1994 P.N.A.S. 91, 12676-12680; Weng and Weiss 1998 J. Virol. 72, 9676-9682). Second, C-34 peptide variants show that the C-34 inhibitory activity depends on its ability to bind to the pocket (Chan et al. 1998 P.N.A.S.

95, 15613-15617). Third, it may be difficult for HIV-1 to become resistant to drugs that target this pocket because the residues comprising this pocket are highly conserved; the segment of mRNA encoding these residues is part of the Rev-response element (Malim et at. 1989 Nature 338, 254-257; Zapp and Green 1989 Nature 342, 714-716); and C-peptides lacking pocket binding residues are more vulnerable to the emergence of resistant virus than C-peptides that contain pocket-binding residues (Rimsky et al. 1998 J. Virol. 72, 986-993). Small molecules that bind to the hydrophobic pocket of the gp41 core might be expected to function as inhibitors in the same dominant negative manner as C-peptides (Chan et al. Cell 89, 263-273). Thus, a strong need exists for the understanding of the structure and sequence of the pocket so one may screen for pharmaceutical compounds which have the ability to bind to the pocket thereby acting as antagonists to the virus.

All U.S. patents and publications referred to herein are hereby incorporated in their entirety by reference.

SUMMARY OF THE INVENTION

The present invention is directed to novel proteins related to the HIV-1 gp41 protein which are useful for screening compounds for anti-HIV activity. In particular, the present invention includes an isolated nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and a nucleotide sequence having at least 65% identity to a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4.

The present invention also includes a purified polypeptide encoded by a nucleotide sequence selected from

the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and a nucleotide sequence having at least 65% identity to a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4.

Additionally, the present invention encompasses a purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity, preferably at least 75% similarity, and more preferably at least 90% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8.

Further, the present invention also includes a vector comprising one of the above nucleotide sequences as well as a host cell comprising the vector.

The present invention also encompasses a method of producing a protein having an unoccupied Trp-Trp-Ile pocket comprising the steps of: a) isolating a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and a nucleotide sequence having at least 65% identity to a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4; b) constructing a vector comprising 1) said nucleotide sequence of step (a) linked to 2) a promoter in an operable manner; and c) transforming a host cell with said vector of step (b) under time and conditions suitable for expression of the protein.

Additionally, the present invention includes a method of detecting a compound which binds to gp41 protein comprising the steps of: a) contacting the compound of interest with a polypeptide having an amino acid sequence

selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8, for a time and under conditions sufficient for the formation of compound/polypeptide complexes; and b) detecting presence of the complexes, wherein detection indicates presence of a compound which binds to gp41 protein. The polypeptide of step (a) may be attached to a solid phase prior to performing step (a). Furthermore, the solid phase may be, for example, a porous or non-porous material, a latex particle, a magnetic particle, a microparticle, a bead, a membrane, a microtiter well or a plastic tube.

The present invention also includes a method of detecting a compound which binds to gp41 protein comprising the steps of: a) adding an indicator reagent capable of generating a measurable signal to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8, for a time and under conditions sufficient for the formation of indicator reagent/polypeptide complexes; b) contacting said indicator reagent/polypeptide complexes with said compound, for a time and under conditions sufficient for the formation of indicator reagent/polypeptide/compound complexes; and c) detecting a measurable signal generated by the indicator reagent, the measurable signal indicating presence of a compound which binds to gp41 protein. The indicator reagent may be, for example, an enzyme such as horseradish

peroxidase, beta-galactosidase or alkaline phosphatase, a luminescent compound, a radioactive element, a visual label or a chemiluminescent compound.

Moreover, the present invention also includes an antibody produced in response to or directed against any of the polypeptides described above. The antibody may be either monoclonal or polyclonal.

The present invention also encompasses a method for producing antibodies to gp41 protein comprising administering to a mammal a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8, in an amount sufficient to produce an immune response.

Additionally, the invention also includes a vaccine for treatment of human immunodeficiency virus type 1 (or Autoimmune Immunodeficiency Syndrome caused thereby) comprising an antibody noted above and a pharmaceutically acceptable excipient.

Also, the present invention encompasses a method of detecting compounds which bind to gp41 protein from a mixture of compounds having unknown binding properties comprising the steps of: a) contacting at least one polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group cons SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:9, with said compound mixture for a time and under conditions sufficient for the formation of polypeptide/compound

complexes; b) passing said mixture through a means having pores which allow only certain sized molecular weigh molecules to pass through; and c) detecting retained polypeptide/compound complexes which did not pass through the pores, wherein compounds present in the complexes bind to gp41 protein. The means may be, for example, a filter.

Additionally, the present invention includes a method of detecting compounds which bind to gp41 protein from a mixture of compounds having unknown binding properties comprising the steps of: a) contacting at least one polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO;7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:9, with said compound mixture, for a time and under conditions sufficient for the formation of polypeptide/compound complexes; b) passing the mixture through a means that will pass molecules or complexes of a larger molecular size faster than those of smaller molecular size; and c) detecting the separated polypeptide/compound complexes which passed through the means more quickly than the smaller molecules or complexes, wherein compounds present in the complexes which passed through said means at the faster rate, bind to gp41 protein. This means may be, for example, a size exclusion resin. (See Dunayevskiy et al., Rapid Communications in Mass Spectrometry, 11:1178-1184 (1997) and Kaur et al., Journal of Protein Chemistry, 16(5):505-511 (1997).)

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the manner in which the present protein was constructed (i.e., annealing reaction, fill in,

ligation and subsequent PCR reaction). The nucleotide sequences of the oligos are shown in the annealing reaction. Each oligo hydrogen bonded with complementary sequences of another oligo (e.g., gp41 Antisense #2 with gp41 Sense #1 and gp41 Antisense #3) as the temperature of the reaction decreased from 95 °C to 25 °C.

Figure 2 represents the isolated nucleotide sequences of the proteins of the present invention. In particular, the nucleotide sequences of the gp41 clones 1 (SEQ ID NO:1), 1a (SEQ ID NO:2), 1b (SEQ ID NO:3) and 1c (SEQ ID NO:4) are shown. The linker codons for each clone, (GGGG)R*, (GDG)R, (GSG)P, and (GSNDG)R (SEQ ID NOS:18 through 21, respectively) are boldfaced. *The R residue is present on the natural gp41 sequence. It was changed to a P residue in the clone 1B sequence.

Figure 3 represents a sequence alignment of pNL4-3 (SEQ ID NO:17, GenBank Accession No. M19921), clone 1 and the consensus HIV gp41 (strain HXB2) amino acid sequences. The wild-type sequence pNL4-3 was rearranged in the context of the re-engineered protein (i.e., C-Helix(linker)N-Helix) for purposes of this alignment.

Figure 4 represents the amino acid sequences of the proteins of the present invention (i.e., Clone #1 = SEQ ID NO:5, Clone #1a(GDG)R = SEQ ID NO:6, Clone #1b(GDG)P = SEQ ID NO:7, Clone #1c(GSNDG)R = SEQ ID NO:8).

Figure 5 represents the 1-D ^{1}H-NMR spectrum of clone 1b and selective shifting of resonances upon the addition of Trp-Arg-Trp-Arg-Ile pentapeptide (SEQ ID NO:28).

Figure 6 shows the crystal structure of Trimethyl Lead Acetate as soaked into the crystals of clone 1B.

Figure 7 represents the results of a sedimentation equilibrium study of the gp41 clone 4 construct.

Figure 8 illustrates the binding of cyclic D-peptides to gp41 examined by centrifugal enhanced affinity selection.

Figure 9 represents the nucleotide sequence of clone 4 (SEQ ID NO:9). The four Gly linker codons (SEQ ID NO:22) are shown in boldface.

Figure 10 represents the nucleotide sequences of gp41 clones 4 (SEQ ID NO:9), 4a (SEQ ID NO:10), 4b (SEQ ID NO:11), and 4c (SEQ ID NO:12). The linker codons are boldfaced. *The R residue is present on the natural gp41 sequence. It was changed to a P residue in the clone 4B sequence.

Figure 11 illustrates the amino acid sequences of clones 4 (SEQ ID NO:13), 4a (SEQ ID NO:14), 4b (SEQ ID NO:15), and 4c (SEQ ID NO:16). The sequence designated as MGHHHHHHHSSGHIDDDDK (SEQ ID NO:27) represents a His tag/E. K. Cleavage site.

DETAILED DESCRIPTION OF THE INVENTION

The subject invention relates to proteins which may be utilized to screen for anti-human immunodeficiency (HIV) compounds. Such compounds may, in turn, be used to treat or prevent HIV-afflicted or susceptible individuals.

The isolated nucleotide sequences which encode the amino acid sequences of the proteins of the present invention are shown in Figure 2. The present invention encompasses not only these nucleotide sequence but fragments thereof, complements of the full sequences thereof, and fragments of the complements. Additionally, the invention includes sequences corresponding to (i.e., having identity to) or complementary to at least about 65%, preferably at least about 75%, and more preferably at least

about 90% of the nucleotide sequences shown in Figure 2. Furthermore, the present invention includes fragments of the full sequences, complements of these sequences, as well as fragments of the complements.

For purposes of the present invention, a "fragment" of a nucleotide sequence is defined as a contiguous sequence of approximately at least 120, preferably at least about 140, more preferably at least about 160 nucleotides, and even more preferably at least about 200 nucleotides corresponding to a region of the specified nucleotide sequence.

Furthermore, for purposes of the present invention, a ''complement'' is defined as a sequence which pairs to a given sequence based upon base-pairing rules. For example, a sequence A-G-T in one nucleotide strand is ''complementary'' to T-C-A in the other strand.

Sequence identity or percent identity is the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm may be extended to use with peptide or protein sequences using the scoring matrix created by Dayhoff, Atlas of Protein Sequences and Structure, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-66763 (1986). An implementation of this algorithm for nucleic acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in the

BestFit utility application. In particular, the default parameters of Gap Creation Penalty of 8 and Gap Extension Penalty of 2, as described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI), were used Other equally suitable programs for calculating the percent identity or similarity between sequences are generally known in the art.

Sequences related to the nucleotide sequence of the present invention may be derived from non-human sources (e.g., bacterial, viral, mammalian, etc.) and are also covered by the present invention. Functional equivalents of the above-sequences (i.e., sequences having the ability to conform to the secondary and tertiary structure of the pocket) are also encompassed by the present invention and hybridize to the present nucleotide sequences.

A nucleic acid molecule is ''hybridizable'' to another nucleic acid molecule when a single-stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and ionic strength (see Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition (1989), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). The conditions of temperature and ionic strength determine the ''stringency'' of the hybridization. ''Hybridization'' requires that two nucleic acid sequences contain complementary sequences. However, depending on the stringency of the hybridization, mismatches between bases may occur. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and

the degree of complementation. Such variables are well known in the art. More specifically, the greater the degree of similarity or homology between two nucleotide sequences, the greater the value of Tm for hybrids of nucleic acids having those sequences. For hybrids of greater than 100 nucleotides in length, equations for calculating Tm have been derived (see Sambrook et al., supra). For hybridization with shorter nucleic acids, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., supra).

Additionally, the present invention encompasses the purified polypeptides or proteins encoded by the nucleotide sequences illustrated in Figure 2. The amino acid sequences of the proteins are shown in Figure 4. The invention also includes those peptides, polypeptides or proteins, or fragments thereof, having an amino acid sequence that has at least about 65% amino acid similarity, preferably at least about 75% amino acid similarity, and more preferably at least about 90% amino acid similarity to the amino acid sequences resulting from the translation of the nucleotide sequences present in Figure 2. For purposes of the present invention, ''similarity'' is defined as the exact amino acid to amino acid comparison of two or more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. ''Percent similarity'' is calculated between the compared polypeptide sequences using programs known in the art (see above).

It should also be noted that the clones described herein have an initiating Met codon (ATG) engineered at the 5' end of the gene. Gp41 proteins purified from *E. coli* harboring these cDNA clones have the N-terminal initiator methionine residue clipped off. This phenomenon occurs during the process of translation within the bacterial host. The formyl groups present on the initiating Met residue of polypeptides in prokaryotes (e.g., *E. coli*) are generally removed rapidly during biosynthesis by a ''deformylase'' enzyme. In many case, the Met residue is also removed by a specific aminopeptidase. Consequently, not all mature proteins have Met as the amino-terminal residue, even though all are synthesized in this manner.

In terms of usage of the proteins of the present invention, such purified proteins may be utilized for many purposes. For example, the proteins may be used to screen for compositions which bind to the ''Trp-Trp-Ile'' pocket of the proteins and would therefore consequently bind to the pocket of gp41. In particular, if one is able to identify compounds that bind to gp41, then one may prevent further replication of the virus and the resulting pathophysiological changes (e.g., wasting, dementia, liver involvement, etc.) associated with AIDS One may also prevent initial infection with the HIV Type I.

The identification of compounds which bind to gp41, by use of the proteins of the present invention, may be carried out by the use of, for example, drug screening assays. Initially, however, one must produce the proteins for use in the assays.

Expression of the Proteins:

In order to express the proteins of the present invention, a vector is first constructed comprising at least one of the isolated DNA sequences encoding the complete protein or proteins of interest. The vector may be, for example, a plasmid, a bacteriophage or a cosmid. The vector is then introduced into a eukaryotic or prokaryotic host cell (e.g., mammalian cells (e.g., Chincese Hamster Ovary (CHO) cells, yeast cells, insect cells and bacterial cells (e.g., E. coli)) under time and conditions suitable for production of the protein, by a method commonly known in the art (see, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press (1989)).

The protein may be isolated from the host cell, expressing the protein, according to procedures known in the art including, for example, ammonium sulfate precipitation, fractionation column chromatography (e.g., ion exchange, gel filtration, electrophoresis and affinity chromatography) and ultimately by crystallization (see, e.g., "Enzyme Purification and Related Techniques", Methods in Enzymology, 22, 233-577 (1971)).

Of course, the proteins of the present invention, or portions thereof, may also be prepared by chemical synthesis by use of techniques well known in the art, such as by solid phase synthesis (Merrifield, 1964, J. Am. Chem. Assoc. 85:2149-2154) or synthesis in homogenous solution (Houbenweyl, 1987, Methods of Organic Chemistry, ed. E. Wansch, Vol. 15 I and II, Thieme, Stuttgart).

Drug Screening Assays

Once the "final" proteins of interest are produced, drug screening assays may be carried out in order to identify antagonists of the present proteins and thus compounds which bind to gp41. One such assay involves adding a signal generating compound or label (e.g., chromogens, radioisotopes (e.g., 125I, 131I, 32P, 3H, 35 S and 34C), fluorescent compounds (e.g., fluorescein and rhodamine), chemiluminescent compounds, particles (visible or fluorescent), nucleic acids, complexing agents, or catalysts such as enzymes wherein addition of a chromo-, fluoro- or lumo-genic substrate results in generation of the detectable signal (e.g., alkaline phosphatase, acid phosphatase, horseradish peroxidase, beta-galactosidase, and ribonuclease)) to the "final protein" comprising the linker and the unnatural domain. The labelled protein is then added to the composition of interest. Upon binding between the protein and the composition of interest, a signal is generated by the signal-generating compound, either visually or through instrumentation. Such a signal indicates that the composition of interest would bind to gp41 in vitro or in vivo.

It should be noted that the protein may be added to a solid phase initially, if desired. Examples of solid phases include porous and non-porous materials, latex particles, magnetic particles, microparticles, beads, membranes, microtiter wells and plastic tubes.

Additionally, in order to identify compositions which bind to gp41, using the present proteins, one may utilize an Affinity-Selection method known to those of ordinary skill in the art (see, e.g., U.S. Patent No. 5,891,742; U.S. Patent No. 5,891,742; U.S. Patent No. 5,670,326). Briefly, one or more of the purified proteins of the

present invention is mixed with several test compounds of interest. The mixture is passed through a filter which will retain protein combined with bound compound and which allows unbound compound to pass through. A typical membrane of use is a 5000-10000 Dalton membrane. Compositions that bind to the protein(s) and, in particular, the Trp-Trp-Ile pocket of the protein(s), will be retained by the filter. The unbound compounds are not retained by the filter and can therefore be separated from the bound compositions. The retained compositions (i.e., those bound to the protein) may be utilized in preventing and treating, for example, AIDS. Alternatively, a sizing column with exclusion limit of 5000 Dalton or below, such as Sephadex G-25, can be employed. A third method for separating bound and unbound molecules is centrifugal enhanced affinity selection (e.g., spin screen (see U.S. Patent Appln. Ser. No. 09/270,427 incorporated in its entirety, by reference, herein).

It should also be noted that the proteins of the present invention may be used in drug screening assays or methods other than Affinity-Selection. For example CrystaLead technolgy may be utilized (see WO 99/45379) as well as SARbyNMR technology (see WO 98/48264) to identity pharmaceutical compounds having the ability to bind to gp41. Drug screening, by use of the proteins of the present invention, may also be accomplished using high-throughput screening (see Eckert et al., WO 00/06599). Computational docking may also be used to screen for pharmaceuticals of interest (see Debnath et al., 1999 J. Med. Chem. 42, 3203-3209 and Kuntz et al., 1994 Acc. Chem. Res. 27(5), 117-123).

Another method which may be utilized for screening compounds which bind to the proteins of the present invention and thus gp41 is termed ''spin screen'' (see, e.g., Holzman et al., Abstract, ''Identification of Ligands by Spin Screen: A Lead Discovery and Refinement Process'', presented at ''Drug Discovery Technology'' seminar, August 16-19, 1999, Boston, MA).

Once compounds have been identified which have the ability to bind to gp41, such compositions may be administered to patients having, for example, Autoimmune Deficiency Syndrome (AIDS) or a precursor stage thereof, or may be utilized in the prevention of the disease. Furthermore, any condition caused by a virus which comprises the gp41 protein, may also be treated by administration of compounds, identified by the proteins and methods of the present invention, which bind to gp41.

The pharmaceutical composition may comprise a therapeutically effective amount of the active drug, discovered above, and an appropriate physiologically acceptable carrier (e.g., water, buffered water or saline). The dosage, form (e.g., suspension, tablet, capsule, etc.), and route of administration of the pharmaceutical composition (e.g., oral, topical, intravenous, subcutaneous, etc.) may be readily determined by a medical practitioner and may depend upon such factors as, for example, the patient's age, weight, immune status, and overall health.

Antibody Production

The present invention also encompasses antibodies (e.g., monoclonal or polyclonal) or portions thereof produced in response to the present proteins. (See U.S. Patent No. 4,196,265 for a discussion of the production of monoclonal antibodies; see also Kohler et al., Nature, 1975, 256:495-497. For a discussion of polyclonal antibody production, see, e.g., ANTIBODIES: A LABORATORY MANUAL, Cold Spring Harbor Laboratory, 1988). These antibodies may be produced, for example, by injecting a mammal (e.g., a mouse or a goat) with the proteins described herein or a portion thereof (see, e.g., ANTIBODIES: A LABORATORY MANUAL, supra; see also U.S. Patent No. 4,196,265). Furthermore, an adjuvant such as Freund's adjuvant may also be injected with the protein. An immune response will then be elicited and antibodies will be produced from the animal. Such antibodies may then be purified from the blood of the mammal.

The anti-"protein" antibodies of the present invention have significant utility in immunoassays for the detection of gp41 due to their cross-reactivity therewith. Thus, such antibodies may be utilized for diagnosis as well as prognostic monitoring of HIV as indicated through the presence of gp41. Immunoassays included within the present invention encompass, but are not limited to, those described in U.S. Pat. No. 4,367, 110 (double monoclonal antibody sandwich assay) and U.S. Pat. No. 4,452,901 (western blot). Other assays include, for example, immunoprecipitation of labeled ligands and immunocytochemistry, both *in vitro* and *in vivo*. Preferred assays are, for example, enzyme linked immunosorbent assays (ELISAs) and radioimmunoassays(RIA), both of which are well known in the art.

Additionally, the present invention includes vaccines comprising the above-described antibodies or portions thereof. The antibodies may be administered to an individual (e.g., an individual recently infected with HIV, an individual having AIDS or a non-infected individual) with, for example, an appropriate carrier (e.g., water, buffered water or saline). Subsequent to administration of the vaccine, the antibodies will bind to expressed gp41 in the body in order to form a complex, thereby preventing further replication of the virus or even initial replication of the virus in a patient recently "infected" such that symptoms would never manifest.

The present invention may be illustrated by the use of the following non-limiting examples:

Example I

Construction of the Nucleotide Sequence Corresponding to the gp41 Gene

The nucleotide sequence corresponding to the reengineered gp41 gene was synthesized as a series of four, overlapping oligonucleotides. The oligonucleotides had the following sequence:

gp41 Sense #1 (SEQ ID NO:23)

5'-TACACAAGCTTGATCCACTCTCTGATCGAAGAAAGCCAGAACCAGCAGGAAAAAAACGAACAGG-3'

gp41 Antisense #2 (SEQ ID NO:24)

5'-CCAGACAGAAGCTGACGACCACCACCACCGTCCAGTTCTAGAAGTTCCTGTTCGTTTTTTTCCTGCTGG-3'

gp41 Sense #3 (SEQ ID NO:25)

5'-GGTCGTCAGCTTCTGTCTGGTATCGTTCAGCAGCAGAACAATCTGCTGCGTGCTATCGAAGCTCAGCAGCATC-3'

gp41 Antisense #4 (SEQ ID NO:26)

5'-TTCAACAGCCAGGATACGAGCCTGAAGCTGTTTGATACCCCAAACGGTCAGTTGCAGCAGATGCTGCTGAGCT
TCGATAGCACG-3'.

These oligonucleotides were phosphorylated with T4 polycucleotide kinase, annealed, and used as a template for T4 DNA polymerase to fill in the gaps (Fig. 1). Specifically, ten pmols of each primer was kinased with bacteriophage T4 polynucleotide Kinase as per the instructions of the manufacturer, Life Technologies (Rockville, MD). One pmol of each kinased primer was mixed together in a 20ul annealing reaction. The final buffer components were 20 mM Tris-HCl pH 7.4, 2 mM $MgCl_2$, 50 mM NaCl. Eppendorf tubes containing these reactions were sealed and inverted in a 2 liter beaker of H_2O heated to 100 degrees C. These reactions remained in the one liter beaker until the water had reached room temperature. After this the DNA was precipitated with ethanol, washed with 70% ethanol and dried. The annealed DNA was dissolved in T4 Polymerase buffer as per the specifications of the manufacturer (Life Technologies, Rockville, MD). T4 DNA polymerase was added to the reaction to fill in the single-stranded gaps after the reaction was placed at 37 degrees C for 1 hour. T4 DNA ligase (Life Technologies, Rockville, MD) was then added to seal the nicks that occurred within the template.

After ligation with T4 DNA Ligase, the resulting DNA molecule was used as a template for PCR to amplify the reengineered sequence for cloning into an appropriate expression vector (e.g., pET 19b or pET 21a). Positive clones were sequenced followed by protein expression and purification studies. Bacteria, BL21[DE3] were induced to

express gp41 by the addition of Isopropyl B-D-Thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM in a 1 liter LB culture which had an $O.D._{600}$ reading of 0.6. The addition of IPTG induced the expression of the T7 RNA polymerase gene which is integrated in the bacterial chromosomal DNA. This DNA-dependent RNA polymerase then specifically recognized the T7 promoter sequence on the recombinant plasmid DNA (e.g., pET21a and pET19b Novagen, Madison, WI) and drove the transcription of the gp41 mRNA. This mRNA was then translated to a high level by the *E. coli* ribosomes.

After expression for two hours at 37 degrees C, the bacteria were harvested by centrifugation, washed with 40 mM Tris-HCl, 100 mM NaCl buffer, collected by centrifugation again and the cell pellet was stored at -80 degrees C. Prior to the addition of IPTG and at the time of harvest, 1 mL aliquots were taken out for analysis on SDS-PAGE to see if there was a protein expressed of the expected molecular weight. Bacterial pellets demonstrating expression of gp41 were lysed by passage through a French pressure cell. This lysate was centrifuged at 10,000 x g to separate the insoluble material (P10) from the soluble material (S10). Ammonium Sulfate was added to the S10 to a final saturation of 35%. The precipitated protein from this treatment was collected at 12,000 rpm. The protein pellet was dissolved in 5 mls of Lysis Buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, and 5% glycerol) and dialyzed against Lysis Buffer for two, 1 liter buffer changes. This protein was then loaded on a Uno Q6 column (Cation Exchanger, BioRad, Hercules, CA) and eluted by a 0.1 M NaCl to 1.0 M NaCl linear gradient. Peak fractions were pooled and dialyzed against 20 mM MES pH 6.0, 100 mM NaCl, 1 mM

EDTA, and 5% glycerol. Precipitated protein was collected by centrifugation and dissolved in 10 mM Bis Tris pH 7.7.

Example II

Linkers Used to Produce Pocket Formation

Linkers of varying length were engineered into the DNA sequence encoding gp41 (see Fig. 2 for nucleotide and 4 peptide sequences). In particular, visual inspection of the proximity of the two termini of the N- and C-helices indicated that peptides of 3, 4, and 5 amino acid length could span the required distance.

The original clone (Clone 1) containing the nucleotides encoding the GGGG linker was designed in the initial oligo annealing fill in experiment (Fig. 1a). Clone 1a, clone 1b, and clone 1c were constructed by standard PCR where the sense primer was long enough (i.e. 117 nucleotide sense primers for clone 1a and clone 1b and a 123 nucleotide sense primer for clone 1c) to cover the linker region. The antisense primers were the same for each clone. After PCR, the amplified DNAs of each linker derivative were cut with restriction endonucleases for cloning into the appropriate expression vector (e.g. pET21a). Plasmids positive for DNA inserts were sequenced to verify the changes in the linker sequenced. These plasmid DNAs were then used to transform E. coli BL21[DE3] for expression and purification as described above.

Example III

1-D ^{1}H-NMR Spectrum of Clone 1b and Selective Shifting of Resonances Upon Addition of Pentapeptide

Nuclear magnetic resonance (NMR) spectroscopy is a proven technique for the detection of ligand binding to proteins. Application of this technique for screening, or structure activity relationships (SAR) by NMR (Shuker, et al., *Science 274*:1531-1534 (1996)) has been successful to identify drug leads against several proteins (WO 97/18471, published May 22, 1997 and WO 97/18469, published May 22, 1997, both of which are incorporated herein by reference). This technique relies on detecting chemical shifts of amide proton and nitrogen atoms resulting from changes in the chemical environment of the peptide backbone, such as those that occur upon ligand binding.

Based on the technique's sensitivity, experiments were designed to evaluate the binding of the peptide with the sequence WRWRI to clone 1B of GP41. The protein encoded by GP41 clone 1b was isotopically labeled with ^{15}N-ammonium chloride in *E. coli* and purified. The protein was concentrated to 0.4 mM in 10 mM sodium phosphate buffer, pH 7.9 in H_2O/D_2O (9:1). Amide proton and nitrogen chemical shift measurements were obtained from ^{15}N heteronuclear single quantum coherence experiments at 45 degrees C on a Bruker DRX500 NMR spectrometer (Bruker Instruments, Karlsruhe, Germany). To detect binding of the WRWRI peptide to GP41, the NMR experiment was repeated in the presence of 0.5 mM of the WRWRI peptide using the same buffer conditions described above. The ^{15}N heteronuclear single quantum coherence spectra for the gp41 protein, in the presence and absence of this peptide, are shown in Figure 5. Both chemical shift changes and line broadening in the presence of the WRWRI peptide indicate that binding of the two molecules had occurred.

Example IV

Crystalline Structure of Clone 1b

The gp41 clone 1B protein was expressed and purified as described above. The protein was concentrated to approximately 7 mg/ml in 20 mM Hepes buffer at pH 7.5. In this solution, the protein was stable for several months at 4 degrees C. The initial crystals were identified from Hampton Crystal Screen I (Hampton Research, 25431 Cabot Rd., Suite 205, Laguna Hills, CA, 92653-5527) from condition number 23 (0.2 M magnesium chloride hexahydrate, 0.1 M Hepes pH 7.5, 30% PEG 400) and from a variety of other related conditions. Crystal growth was optimized with hanging drop vapor diffusion at 17 degrees C under the conditions of 0.2 M magnesium chloride hexahydrate, 0.1 M Tris pH 8.5 and screening 33-40% PEG 400 at a protein concentration of 7 mg/ml. Crystals grow to an average size of 400 X 400 X 200 microns in approximately 1-2 weeks.

The crystals could be frozen at 150 degrees Kelvin for data collection, using an oxford cryo-system by soaking the crystals for at least 1 hour in 23% PEG 400, 0.15 M magnesium chloride hexahydrate, 0.08 M Tris pH 8.5 and 10% glycerol. Under these conditions the crystals diffract to 1.9 Angstrom resolution at a synchrotron radiation source. Native data to 2.8 Ang. were collected and the crystals belong to the space group P321, with the unit cell parameters a=b=96.9, c=72.7, and alpha=beta=90 and gamma= 120. The structure was solved by molecular replacement using a molecular model based on the coordinates derived from Protein Data Band entry 1ENV and using the molecular replacement program, AMORE (Collaborative Computational

Project, Number 4, 1994. "The CCP4 Suite: Programs for Protein Crystallography". Acta Cryst. D50, 760-763.). The structures were refined using CNX (A. T. Brunger, "The Free R Value: a Novel Statistical Quantity for Assessing the Accuracy of Crystal Structures, Nature 355, 472-474 (1992)) using a combination of simulated annealing maximum likelihood refinement and individual B-factor refinement. Electron density maps were inspected on a Silicon Graphics INDIGO2 workstation using the program package QUANTA 98 (Molecular Simulations Inc., San Diego, CA).

The native structure contains four gp41 clone 1B monomers (one monomer consists of one N-helix/C-helix pair) in the asymmetric unit. Three monomers make one native like gp41 trimer, the other monomer sits on the crystallographic three fold rotation axis which generates a crystallographic trimer. The native structure was refined to 2.8 Ang. resolution with an R= 28% and Rfree of 34%.

Crystals of gp41 could then be soaked with 100 mM trimethyl lead acetate for 1-2 days in 36% PEG 400, 0.2 M magnesium chloride hexahydrate, 0.1 M bis-tris propane, to obtain a lead heavy atom derivative. In the Fo-Fc maps calculated after rigid-body refinement, seven trimethyl lead acetate sites could be identified. One of these sites was centrally located in the target site normally occupied by the Trp Trp Ile of the native C-peptide. Thus, the above data indicates that compounds can be soaked into the binding site.

Example V

Sedimentation Equilibrium Study of gp41 Clone 4 Construct

The gp41 clone 4 protein (see amino acid sequence below) was analyzed at 15K, 20K, 25K, and 30K rpm at 8 °C

in 10mM PO_4, 120mM NaCl, 2.7mM KCl, pH 7.4. (See Figure 9 for the nucleotide sequence of the clone.) Data from the different speeds were analyzed globally by a nonlinear least squares curve fitting of radial concentration profiles using the Marquardt-Levenberg algorithm as implemented in Origin 5.0 (Microcal Software, Inc., Northamptom, MA). A user defined function describing sedimentation behavior of discrete particles was used (see equation 1 of Holzman et al., in "Modern Analytical Ultracentrifugation", Birkhauser, Boston, MA 1994, pp.298-314). Baselines and fixed radius signal values for each data set were allowed to vary independently; however, the molecular weight was held as a global parameter. The partial specific volume of the protein was calculated from the amino acid composition by the method of Cohen and Edsall as 0.729cm^3/gm (Cohen and Edsall, in "Proteins, Amino Acids, and Peptides as Ions and Dipolar Ions", Rheinhold, New York, 1943, Chapter 4, page 157). Buffer density was measured at 8 °C at 1.01456 m/cm^3 in a Mettler-KEM Da-310 density meter (Mettler, Highstown, New Jersey). Closed circles (see Figure 7) are the absorbance data points. The fitted data, and the open circles are the residuals. Initial concentrations were 0.65 mg/ml.

The molecular weight determined from this analysis was 33,000 +/-280, in good agreement with the expected molecular weight for a trimer. This strongly indicates that a construct having a portion of the C-terminal helix of gp-41, followed by a linker, followed by the N-terminal helix, is capable of forming the expected quaternary structure in a manner analagous to the intact protein.

Clone 4 amino acid sequence:

MGHHHHHHSSGHIDDDDKYTSLIHSLIEESQNQQEKNEQELLELDGGGGRQLLS
GIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE

Example VI

Binding of Cylic D-Peptides to gp41 by Centrifugal Enhanced Affinity Selection

The peptide (D10-P1-2K, amino acid sequence Ac-KKGACEARHEWAWLCAA-NH_2, 244 μM) and protein (D10-P5-2K, Ac-KKGACELLGWEWAWLCAA-NH_2, 200 μM) were mixed in the centrifuge tube, vortexed, and spun for 2 hours at an average g-force of 309,880 (1000,000 rpm in a TLA-100 rotor, Beckman, Fullerton, CA). Five fractions, 20 microliters each, were removed by carefully pipetting from under the meniscus and examined by reverse-phase high pressure liquid chromatography. The amount of peptide is followed by peak area of the relevant peaks on the chromatograms. The relevant peak area in each fraction was divided by the sum of the peak areas in all 5 fractions, and the result is plotted versus the additive volume of each fraction. (Figure 8A represents the binding of D10-P1-2K, and Figure 8B represents the binding of D10-P5-2K.) The initial concentrations of protein and peptide were 244 $\mu$and 200 μM, respectively. A closed square indicates the amount of peptide in each fraction after centrifugation in the presence of gp41 clone 1b. Open circles indicate the amount of peptide in each fraction after centrifugation in the absence of gp41.

As evidenced by differential movement of the peptides in the presence of the protein, both peptides bind very well to gp41 clone 1b. As evidenced by the lack of movement in the absence of the protein, both appear to be

quite soluble under the conditions examined. No difference in affinity for gp41 clone 1b was observable from this assay. Recovery of peptides in the presence and absence of protein was good, indicating no loss of ligand during the experiment.

Further, the above results indicate that the present gp41 construct has the ability to bind to a peptide which, in turn, has been shown to bind to a region of the N-terminal helix contained in the present construct.

CLAIMS:

1. An isolated nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and a nucleotide sequence having at least 65% identity to a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4.

2. A purified polypeptide encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, and a nucleotide sequence having at least 65% identity to a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQUENCE ID NO:3 and SEQUENCE ID NO:4.

3. A purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8.

4. A vector comprising said isolated nucleotide sequence of claim 1.

5. A host cell comprising said vector of claim 6.

6. A method of producing a protein having an unoccupied Trp-Trp-Ile pocket comprising the steps of:

a) isolating a nucleotide sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID:4, and a nucleotide sequence having at least 65% identity to a sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4;

b) constructing a vector comprising 1) said nucleotide sequence of step (a) linked to 2) a promoter in an operable manner;

c) transforming a host cell with said vector of step (b) under time and conditions suitable for expression of said protein.

7. A method of detecting a compound which binds to gp41 protein comprising the steps of:

a) contacting said compound with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:8, for a time and under conditions sufficient for the formation of compound/polypeptide complexes; and

b) detecting presence of said complexes, wherein detection indicates presence of a compound which binds to gp41 protein.

8. An antibody directed against said polypeptide of claim 3.

9. A method of detecting a compound which binds to gp41 protein comprising the steps of:

a) adding an indicator reagent capable of generating a measurable signal to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6,SEQ ID NO:7 and SEQ ID NO:8, for a time and under conditions sufficient for the formation of indicator reagent/polypeptide complexes;

b) contacting said indicator reagent/polypeptide complexes with said compound, for a time and under conditions sufficient for the formation of indicator reagent/polypeptide/compound complexes; and

c) detecting a measurable signal generated by said indicator reagent, said measurable signal indicating presence of a compound which binds to gp41 protein.

10. A method of detecting compounds which bind to gp41 protein from a mixture of compounds having unknown binding properties comprising the steps of:

a) contacting at least one polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, and an amino acid sequence having at least 65% similarity to a sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 and SEQ ID NO:9, with said compound mixture for a time and under conditions sufficient for the formation of polypeptide/compound complexes;

b) passing said mixture through a means having pores which allow only certain sized molecular weight molecules to pass through; and

c) detecting retained polypeptide/compound complexes which did not pass through said pores, wherein compounds present in said complexes bind to gp41 protein.

gp41 C-Helix (Gly)$_4$ N-Helix Gene Construction

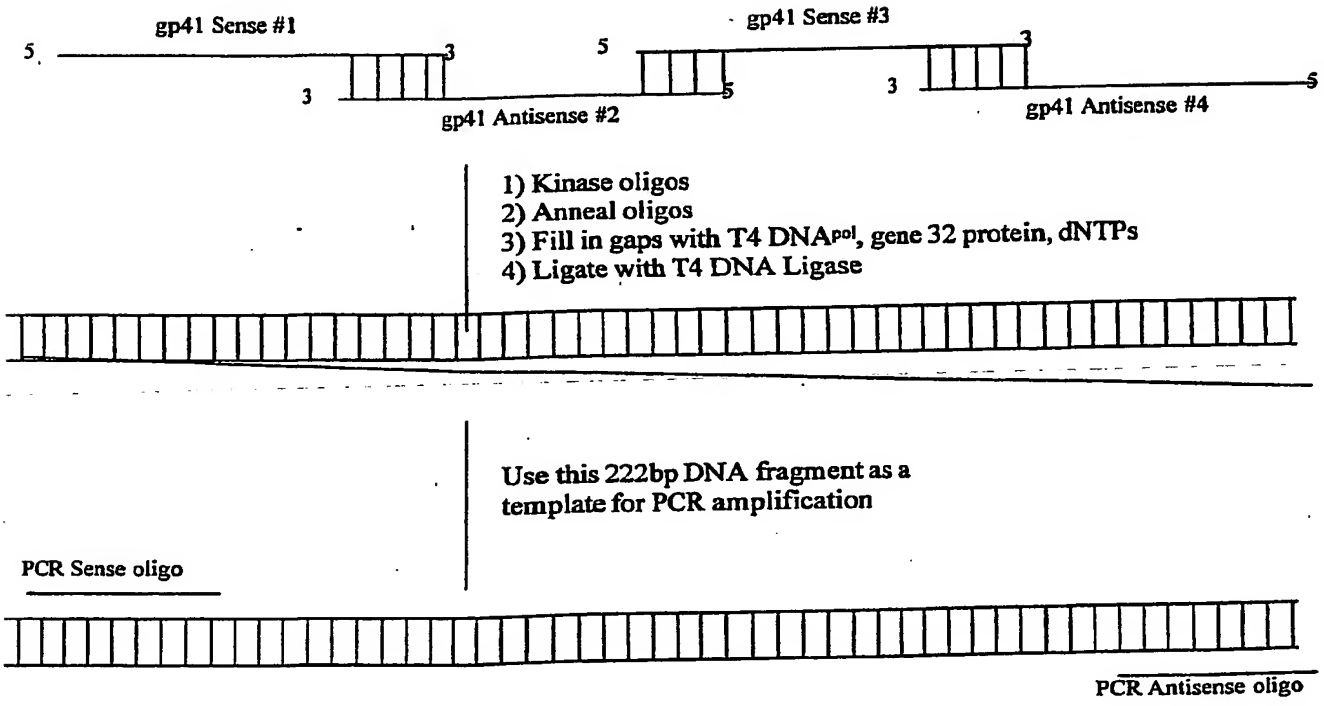


FIGURE 1 (1 OF 2)

Annealing of the gp41 oligos

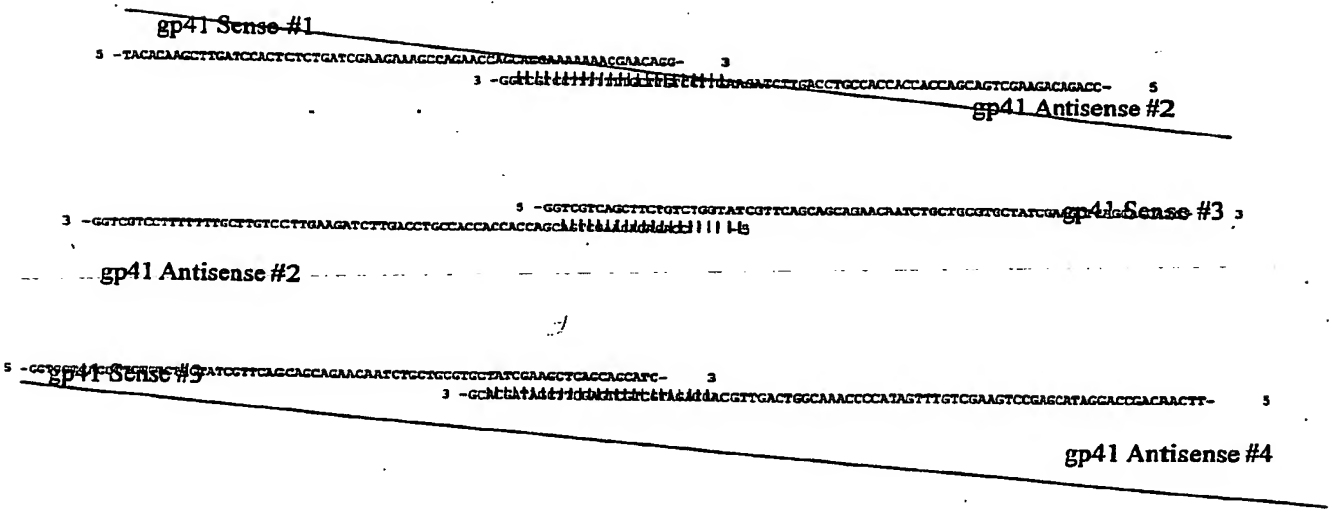


FIGURE 1 (2 OF 2)

gp41 Clone #1 (GGGG)R* Linker

5'- ATG ACA AGC TTG ATC CAC TCT CTG ATC GAA GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA CTG GAC **GGT GGT GGT GGT CGT** CAG CTT CTG TCT GGT ATC GTT CAG CAG CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT CTG CTG CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT ATC CTG GCT GTT GAA -3'

gp41 Clone #1a (GDG)R Linker

5'- ATG ACA AGC TTG ATC CAC TCT CTG ATC GAA GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA CTG GAC **GGT GAC GGT CGT** CAG CTT CTG TCT GGT ATC GTT CAG CAG CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT CTG CTG CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT ATC CTG GCT GTT GAA -3'

gp41 Clone #1b (GDG)P Linker

5'- ATG ACA AGC TTG ATC CAC TCT CTG ATC GAA GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA CTG GAC **GGT GAC GGT CCG** CAG CTT CTG TCT GGT ATC GTT CAG CAG CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT CTG CTG CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT ATC CTG GCT GTT GAA -3'

gp41 Clone #1c (GSNDG)R Linker

5'- ATG ACA AGC TTG ATC CAC TCT CTG ATC GAA GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA CTG GAC **GGT TCT AAC GAC GGT CGT** CAG CTT CTG TCT GGT ATC GTT CAG CAG CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT CTG CTG CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT ATC CTG GCT GTT GAA -3'

FIGURE 2

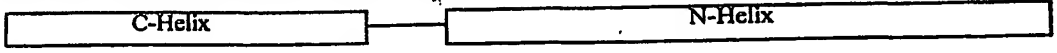
gp41 Sequence Alignment

```
pNL4-3  YTSLIHSLIEESQNQQEKNEQELLELD/     /RQLLSDIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE
C1      M-------------------------(GGGG)-----G-----------------------------------
Cons.   --A--YT-L-----------------------------G-------------------------------V----
```

FIGURE 3

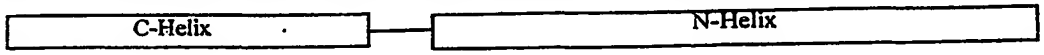
Clone #1

MTSLIHSLIEESQNQQEKNEQELLELD(GGGG)RQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE



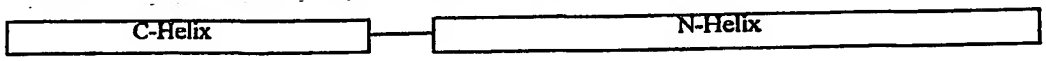
Clone #1a (GDG)R

MTSLIHSLIEESQNQQEKNEQELLELD(GDG)RQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE



Clone #1b (GDG)P

MTSLIHSLIEESQNQQEKNEQELLELD(GDG)PQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE



Clone #1c (GSNDG)R

MTSLIHSLIEESQNQQEKNEQELLELD(GSNDG)RQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE

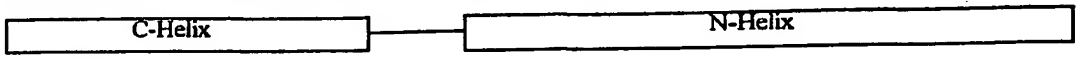


FIGURE 4

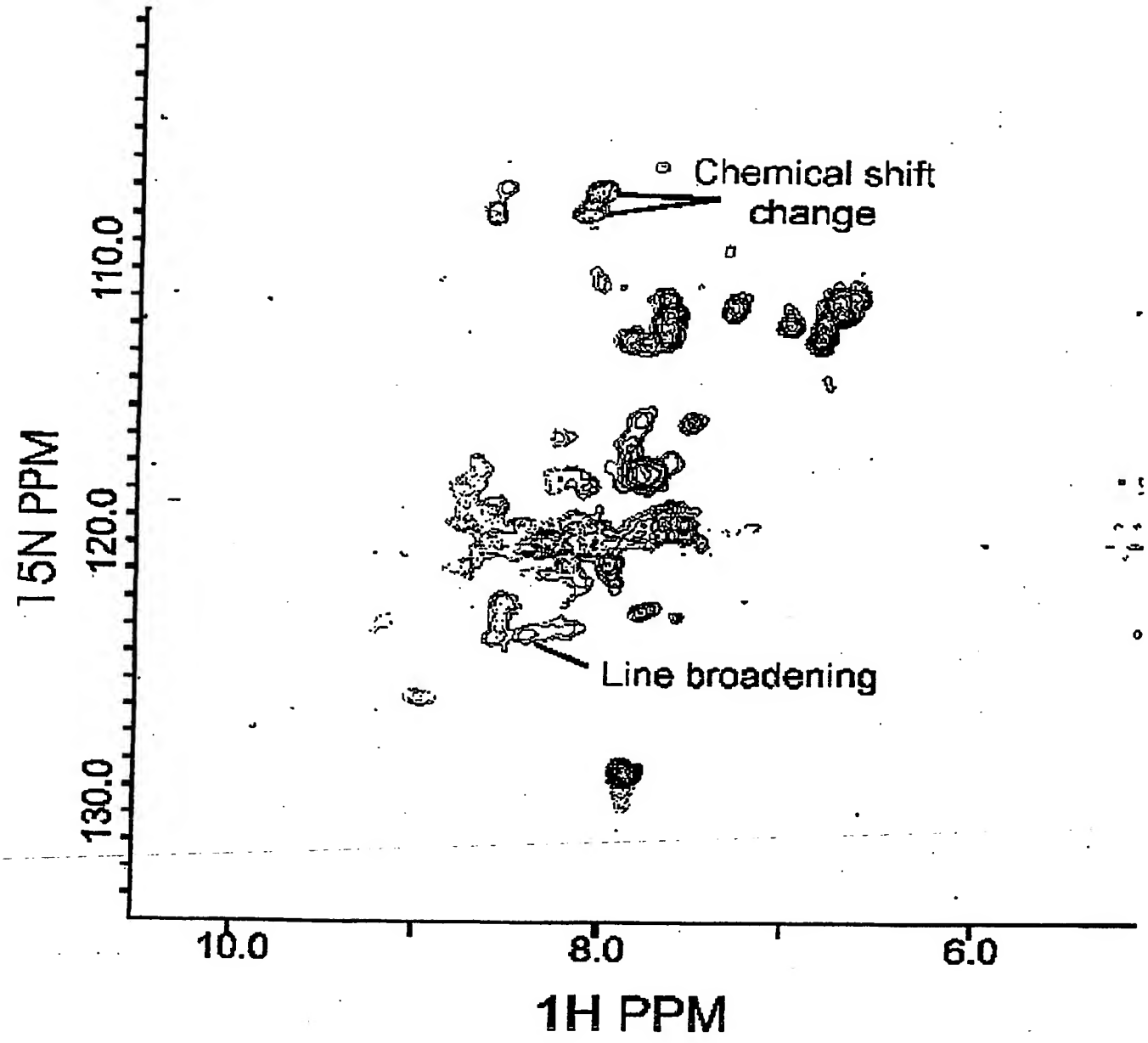

Chemical shift change
Line broadening
15N PPM
110.0
120.0
130.0
10.0
8.0
6.0
1H PPM

FIGURE 5

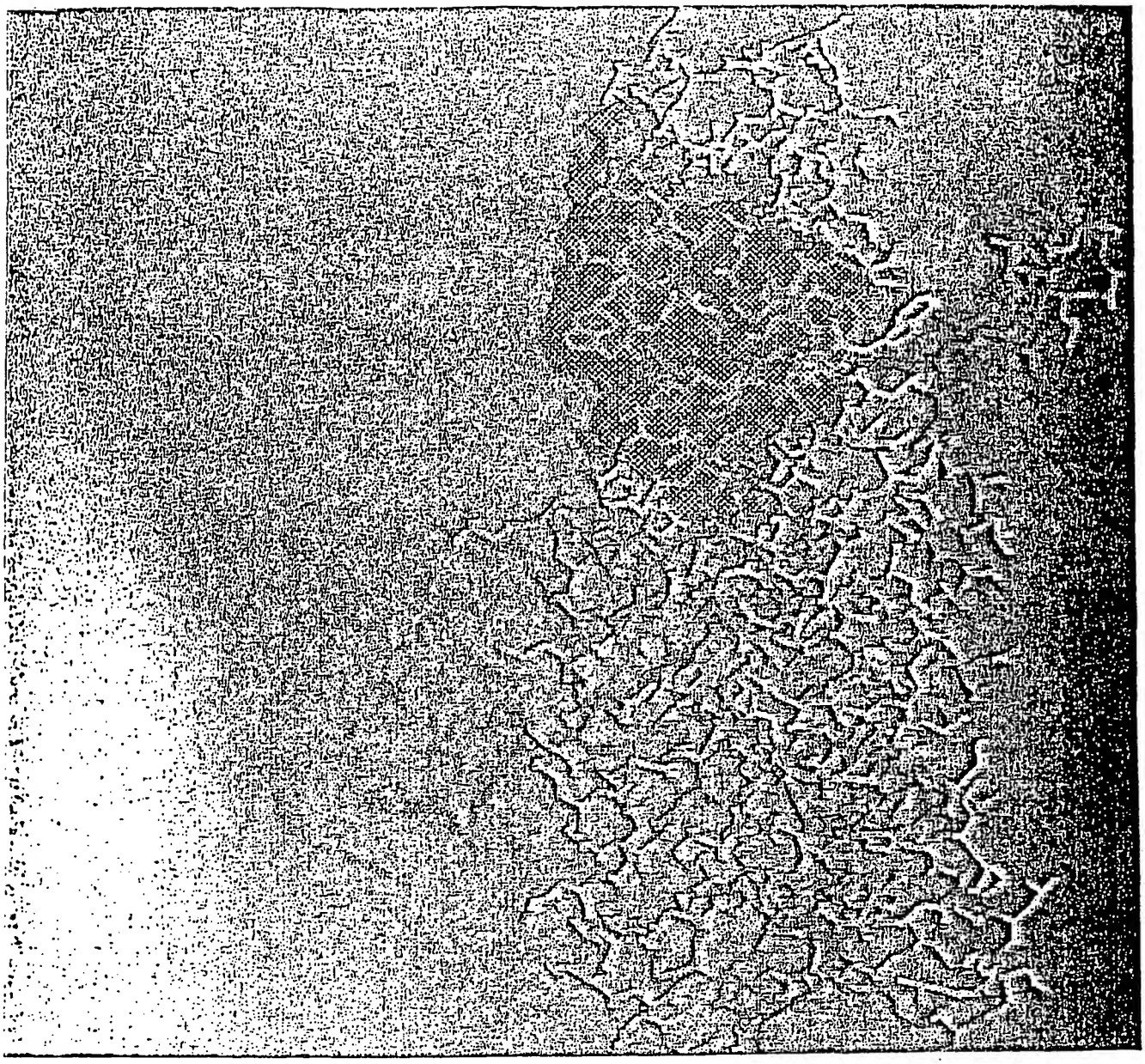


FIGURE 6

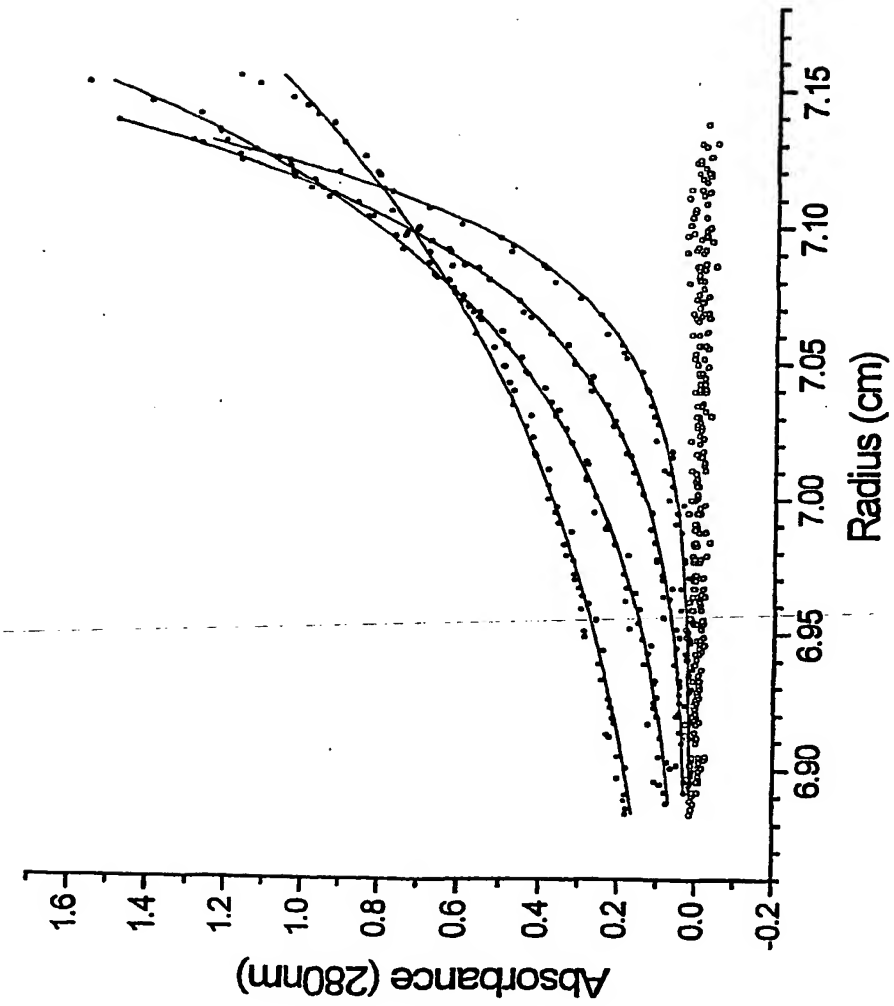
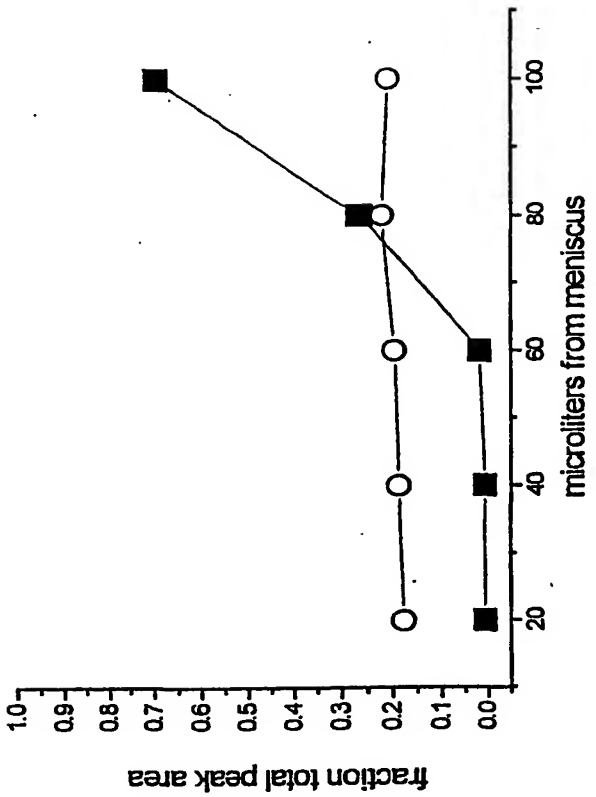
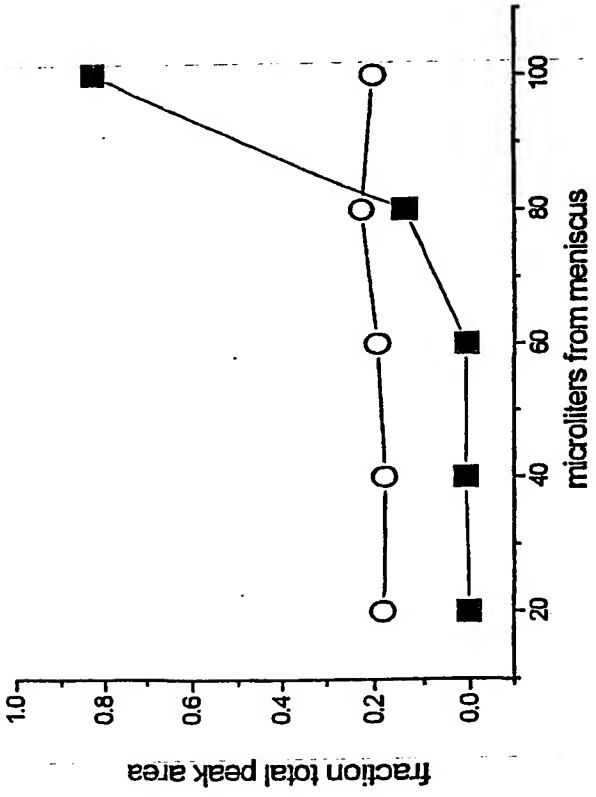

Absorbance (280nm)
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
6.90
6.95
7.00
7.05
7.10
7.15
Radius (cm)

FIGURE 7


fraction total peak area
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
20
40
60
80
100
microliters from meniscus

B.


fraction total peak area
1.0
0.8
0.6
0.4
0.2
0.0
20
40
60
80
100
microliters from meniscus

A.

FIGURE 8

gp41 Clone #4 (GGGG) Linker

5'- ATG GGC CAT CAT CAT CAT CAT CAT CAC AGC AGC GGC CAT ATC
GAC GAC GAC GAC AAG TAC ACA AGC TTG ATC CAC TCT CTG ATC GAA
GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA
CTG GAC GGT GGT GGT GGT CGT CAG CTT CTG TCT GGT ATC GTT CAG
CAG CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT CTG
CTG CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT ATC
CTG GCT GTT GAA -3'

FIGURE 9

gp41 Clone #4 (GGGG)R* Linker

5'- ATG GGC CAT CAT CAT CAT CAT CAT CAC AGC AGC GGC CAT ATC
GAC GAC GAC GAC AAG TAC ACA AGC TTG ATC CAC TCT CTG ATC GAA
GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA
CTG GAC **GGT GGT GGT GGT CGT** CAG CTT CTG TCT GGT ATC GTT CAG
CAG CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT CTG
CTG CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT ATC
CTG GCT GTT GAA -3'

gp41 Clone #4a (GDG)R Linker

5'- ATG GGC CAT CAT CAT CAT CAT CAT CAC AGC AGC GGC CAT ATC
GAC GAC GAC GAC AAG TAC ACA AGC TTG ATC CAC TCT CTG ATC GAA
GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA
CTG GAC **GGT GAC GGT CGT CAG** CTT CTG TCT GGT ATC GTT CAG CAG
CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT CTG CTG
CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT ATC CTG
GCT GTT GAA -3'

gp41 Clone #4b (GDG)P Linker

5'- ATG GGC CAT CAT CAT CAT CAT CAT CAC AGC AGC GGC CAT ATC
GAC GAC GAC GAC AAG TAC ACA AGC TTG ATC CAC TCT CTG ATC GAA
GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA
CTG GAC **GGT GAC GGT CCG CAG** CTT CTG TCT GGT ATC GTT CAG CAG
CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT CTG CTG
CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT ATC CTG
GCT GTT GAA -3'

gp41 Clone #4c (GSNDG)R Linker

5'- ATG GGC CAT CAT CAT CAT CAT CAT CAC AGC AGC GGC CAT ATC
GAC GAC GAC GAC AAG TAC ACA AGC TTG ATC CAC TCT CTG ATC GAA
GAA AGC CAG AAC CAG CAG GAA AAA AAC GAA CAG GAA CTT CTA GAA
CTG GAC **GGT TCT AAC GAC GGT CGT** CAG CTT CTG TCT GGT ATC GTT
CAG CAG CAG AAC AAT CTG CTG CGT GCT ATC GAA GCT CAG CAG CAT
CTG CTG CAA CTG ACC GTT TGG GGT ATC AAA CAG CTT CAG GCT CGT
ATC CTG GCT GTT GAA -3'

FIGURE 10

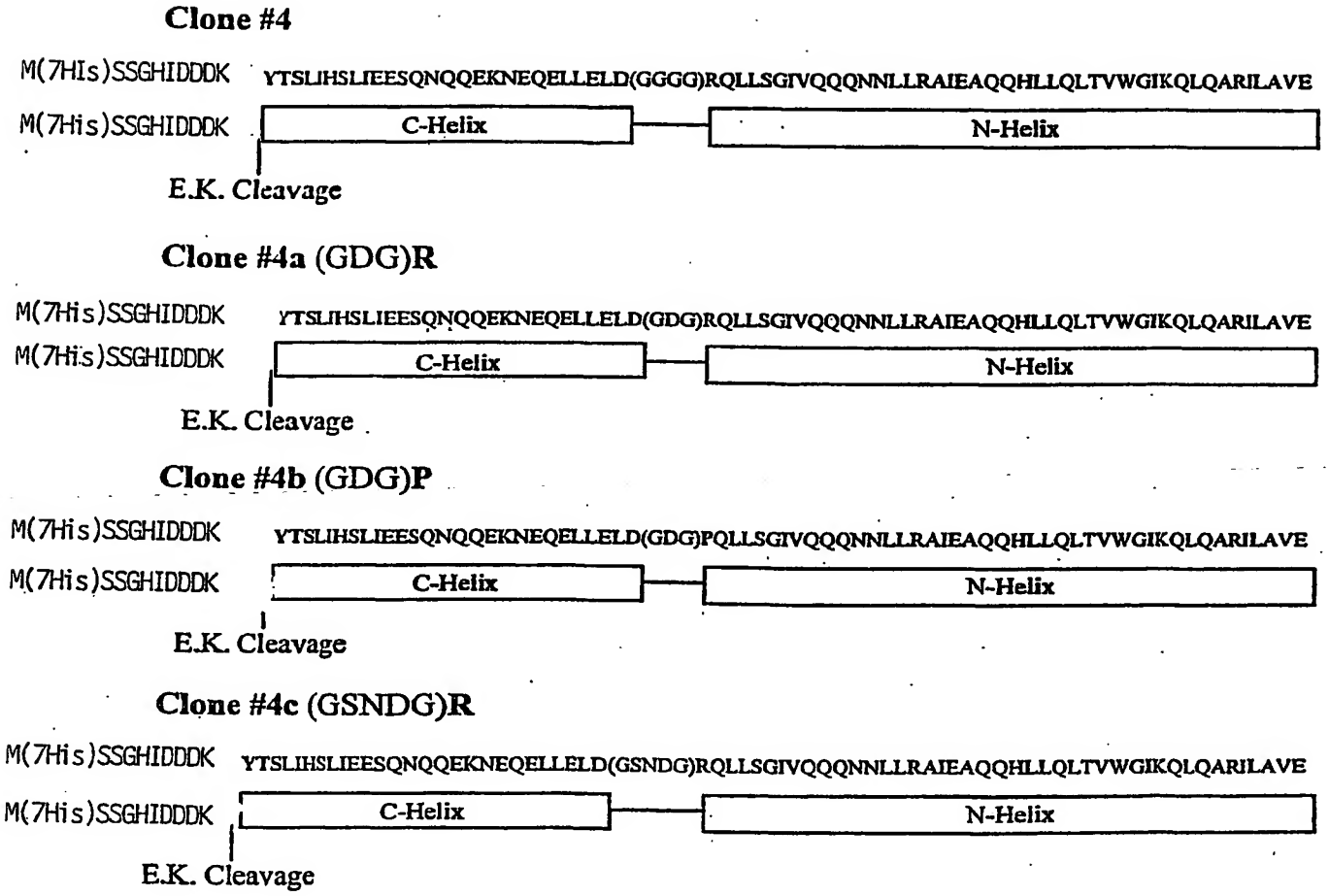

Clone #4
M(7HIs)SSGHIDDDK YTSLIHSLIEESQNQQEKNEQELLELD(GGGG)RQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE
M(7His)SSGHIDDDK
C-Helix
N-Helix
E.K. Cleavage
Clone #4a (GDG)R
M(7His)SSGHIDDDK YTSLIHSLIEESQNQQEKNEQELLELD(GDG)RQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE
M(7His)SSGHIDDDK
C-Helix
N-Helix
E.K. Cleavage
Clone #4b (GDG)P
M(7His)SSGHIDDDK YTSLIHSLIEESQNQQEKNEQELLELD(GDG)PQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE
M(7His)SSGHIDDDK
C-Helix
N-Helix
E.K. Cleavage
Clone #4c (GSNDG)R
M(7His)SSGHIDDDK YTSLIHSLIEESQNQQEKNEQELLELD(GSNDG)RQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVE
M(7His)SSGHIDDDK
C-Helix
N-Helix
E.K. Cleavage

FIGURE 11

SEQUENCE LISTING

<110> Abbott Laboratories
Stewart, Kent D.
Steffy, Kevin R.
Kempf, Dale J.
Harris, Kevin S.
Huth, Jeffrey R.
Stoll, Vincent S.
Harlan, John E.
Ng, Iok C.
Betz, Stephen F.

<120> ENGINEERED CHIMERA OF PROTEIN FRAGMENTS
AND METHODS OF USE THEREOF

<130> 6749.US.01

<140> 09/698,311
<141> 2000-10-27

<160> 28

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 222
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Clone #1

<400> 1
atgacaagct tgatccactc tctgatcgaa gaaagccaga accagcagga aaaaaacgaa 60
caggaacttc tagaactgga cggtggtggt ggtcgtcagc ttctgtctgg tatcgttcag 120
cagcagaaca atctgctgcg tgctatcgaa gctcagcagc atctgctgca actgaccgtt 180
tggggtatca aacagcttca ggctcgtatc ctggctgttg aa 222

<210> 2
<211> 219
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Clone #1a

<400> 2
atgacaagct tgatccactc tctgatcgaa gaaagccaga accagcagga aaaaaacgaa 60
caggaacttc tagaactgga cggtgacggt cgtcagcttc tgtctggtat cgttcagcag 120
cagaacaatc tgctgcgtgc tatcgaagct cagcagcatc tgctgcaact gaccgtttgg 180
ggtatcaaac agcttcaggc tcgtatcctg gctgttgaa 219

<210> 3
<211> 219

<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Clone #1b

<400> 3
atgacaagct tgatccactc tctgatcgaa gaaagccaga accagcagga aaaaaacgaa 60
caggaacttc tagaactgga cggtgacggt ccgcagcttc tgtctggtat cgttcagcag 120
cagaacaatc tgctgcgtgc tatcgaagct cagcagcatc tgctgcaact gaccgtttgg 180
ggtatcaaac agcttcaggc tcgtatcctg gctgttgaa 219

<210> 4
<211> 225
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Clone #1c

<400> 4
atgacaagct tgatccactc tctgatcgaa gaaagccaga accagcagga aaaaaacgaa 60
caggaacttc tagaactgga cggttctaac gacggtcgtc agcttctgtc tggtatcgtt 120
cagcagcaga acaatctgct gcgtgctatc gaagctcagc agcatctgct gcaactgacc 180
gtttggggta tcaaacagct tcaggctcgt atcctggctg ttgaa 225

<210> 5
<211> 74
<212> PRT
<213> Artificial Sequence

<220>
<223> Chimera protein encoded by gp41 Clone #1

<400> 5
Met Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln
1 5 10 15
Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Gly Gly Gly Gly Arg
20 25 30
Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala
35 40 45
Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys
50 55 60
Gln Leu Gln Ala Arg Ile Leu Ala Val Glu
65 70

<210> 6
<211> 73
<212> PRT
<213> Artificial Sequence

<220>
<223> Chimera protein encoded by gp41 Clone #1a

<400> 6
Met Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln
1 5 10 15

Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Gly Asp Gly Arg Gln
20 25 30
Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile
35 40 45
Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln
50 55 60
Leu Gln Ala Arg Ile Leu Ala Val Glu
65 70

<210> 7
<211> 73
<212> PRT
<213> Artificial Sequence

<220>
<223> Chimera protein encoded by gp41 Clone #1b

<400> 7
Met Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln
1 5 10 15
Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Gly Asp Gly Pro Gln
20 25 30
Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile
35 40 45
Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln
50 55 60
Leu Gln Ala Arg Ile Leu Ala Val Glu
65 70

<210> 8
<211> 75
<212> PRT
<213> Artificial Sequence

<220>
<223> Chimera protein encoded by gp41 Clone #1c

<400> 8
Met Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln
1 5 10 15
Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Gly Ser Asn Asp Gly
20 25 30
Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu Arg
35 40 45
Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile
50 55 60
Lys Gln Leu Gln Ala Arg Ile Leu Ala Val Glu
65 70 75

<210> 9
<211> 279
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Clone #4

<400> 9
atgggccatc atcatcatca tcatcacagc agcggccata tcgacgacga cgacaagtac 60
acaagcttga tccactctct gatcgaagaa agccagaacc agcaggaaaa aaacgaacag 120
gaacttctag aactggacgg tggtggtggt cgtcagcttc tgtctggtat cgttcagcag 180
cagaacaatc tgctgcgtgc tatcgaagct cagcagcatc tgctgcaact gaccgtttgg 240
ggtatcaaac agcttcaggc tcgtatcctg gctgttgaa 279

<210> 10
<211> 276
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Clone #4a

<400> 10
atgggccatc atcatcatca tcatcacagc agcggccata tcgacgacga cgacaagtac 60
acaagcttga tccactctct gatcgaagaa agccagaacc agcaggaaaa aaacgaacag 120
gaacttctag aactggacgg tgacggtcgt cagcttctgt ctggtatcgt tcagcagcag 180
aacaatctgc tgcgtgctat cgaagctcag cagcatctgc tgcaactgac cgtttggggt 240
atcaaacagc ttcaggctcg tatcctggct gttgaa 276

<210> 11
<211> 276
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Clone #4b

<400> 11
atgggccatc atcatcatca tcatcacagc agcggccata tcgacgacga cgacaagtac 60
acaagcttga tccactctct gatcgaagaa agccagaacc agcaggaaaa aaacgaacag 120
gaacttctag aactggacgg tgacggtccg cagcttctgt ctggtatcgt tcagcagcag 180
aacaatctgc tgcgtgctat cgaagctcag cagcatctgc tgcaactgac cgtttggggt 240
atcaaacagc ttcaggctcg tatcctggct gttgaa 276

<210> 12
<211> 282
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Clone #4c

<400> 12
atgggccatc atcatcatca tcatcacagc agcggccata tcgacgacga cgacaagtac 60
acaagcttga tccactctct gatcgaagaa agccagaacc agcaggaaaa aaacgaacag 120
gaacttctag aactggacgg ttctaacgac ggtcgtcagc ttctgtctgg tatcgttcag 180
cagcagaaca atctgctgcg tgctatcgaa gctcagcagc atctgctgca actgaccgtt 240
tggggtatca aacagcttca ggctcgtatc ctggctgttg aa 282

<210> 13
<211> 93
<212> PRT
<213> Artificial Sequence

<220>
<223> Chimera protein encoded by gp41 Clone #4

<400> 13
Met Gly His His His His His His His Ser Ser Gly His Ile Asp Asp
1 5 10 15
Asp Asp Lys Tyr Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln
20 25 30
Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Gly Gly
35 40 45
Gly Gly Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu
50 55 60
Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp
65 70 75 80
Gly Ile Lys Gln Leu Gln Ala Arg Ile Leu Ala Val Glu
85 90

<210> 14
<211> 92
<212> PRT
<213> Artificial Sequence

<220>
<223> Chimera protein encoded by gp41 Clone #4a

<400> 14
Met Gly His His His His His His His Ser Ser Gly His Ile Asp Asp
1 5 10 15
Asp Asp Lys Tyr Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln
20 25 30
Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Gly Asp
35 40 45
Gly Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu
50 55 60
Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly
65 70 75 80
Ile Lys Gln Leu Gln Ala Arg Ile Leu Ala Val Glu
85 90

<210> 15
<211> 92
<212> PRT
<213> Artificial Sequence

<220>
<223> Chimera protein encoded by gp41 Clone #4b

<400> 15
Met Gly His His His His His His His Ser Ser Gly His Ile Asp Asp
1 5 10 15
Asp Asp Lys Tyr Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln
20 25 30
Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Gly Asp
35 40 45
Gly Pro Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn Leu Leu
50 55 60
Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val Trp Gly

65 70 75 80
Ile Lys Gln Leu Gln Ala Arg Ile Leu Ala Val Glu
85 90

<210> 16
<211> 94
<212> PRT
<213> Artificial Sequence

<220>
<223> Chimera protein encoded by gp41 Clone #4c

<400> 16
Met Gly His His His His His His His Ser Ser Gly His Ile Asp Asp
1 5 10 15
Asp Asp Lys Tyr Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln
20 25 30
Asn Gln Gln Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Gly Ser
35 40 45
Asn Asp Gly Arg Gln Leu Leu Ser Gly Ile Val Gln Gln Gln Asn Asn
50 55 60
Leu Leu Arg Ala Ile Glu Ala Gln Gln His Leu Leu Gln Leu Thr Val
65 70 75 80
Trp Gly Ile Lys Gln Leu Gln Ala Arg Ile Leu Ala Val Glu
85 90

<210> 17
<211> 70
<212> PRT
<213> Unknown

<220>
<223> Chimera protein of Wild Type HIV-1

<400> 17
Tyr Thr Ser Leu Ile His Ser Leu Ile Glu Glu Ser Gln Asn Gln Gln
1 5 10 15
Glu Lys Asn Glu Gln Glu Leu Leu Glu Leu Asp Arg Gln Leu Leu Ser
20 25 30
Asp Ile Val Gln Gln Gln Asn Asn Leu Leu Arg Ala Ile Glu Ala Gln
35 40 45
Gln His Leu Leu Gln Leu Thr Val Trp Gly Ile Lys Gln Leu Gln Ala
50 55 60
Arg Ile Leu Ala Val Glu
65 70

<210> 18
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker

<400> 18
Gly Gly Gly Gly Arg
1 5

<210> 19
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker

<400> 19
Gly Asp Gly Arg
1

<210> 20
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker

<400> 20
Gly Asp Gly Pro
1

<210> 21
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker

<400> 21
Gly Ser Asn Asp Gly Arg
1 5

<210> 22
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker

<400> 22
Gly Gly Gly Gly
1

<210> 23
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Sense oligonucleotide #1

<400> 23
tacacaagct tgatccactc tctgatcgaa gaaagccaga accagcagga aaaaaacgaa 60
cagg 64

<210> 24
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Antisense oligonucleotide #2

<400> 24
ccagacagaa gctgacgacc accaccaccg tccagttcta gaagttcctg ttcgtttttt 60
tcctgctgg 69

<210> 25
<211> 73
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Sense oligonucleotide #3

<400> 25
ggtcgtcagc ttctgtctgg tatcgttcag cagcagaaca atctgctgcg tgctatcgaa 60
gctcagcagc atc 73

<210> 26
<211> 84
<212> DNA
<213> Artificial Sequence

<220>
<223> gp41 Antisense oligonucleotide #4

<400> 26
ttcaacagcc aggatacgag cctgaagctg tttgataccc caaacggtca gttgcagcag 60
atgctgctga gcttcgatag cacg 84

<210> 27
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> His tag/E. K. Cleavage site

<400> 27
Met Gly His His His His His His Ser Ser Gly His Ile Asp Asp
1 5 10 15
Asp Asp Lys

<210> 28
<211> 5
<212> PRT

<213> Artificial Sequence

<220>
<223> Pentapeptide

<400> 28
Trp Arg Trp Arg Ile
1 5

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 May 2002 (02.05.2002)

PCT

(10) International Publication Number
WO 02/034909 A3

(51) International Patent Classification[7]: C12N 15/11, C07K 14/16, G01N 33/569

(21) International Application Number: PCT/US01/48040

(22) International Filing Date: 26 October 2001 (26.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/698,311 27 October 2000 (27.10.2000) US

(71) Applicant: **ABBOTT LABORATORIES** [US/US]; 100 Abbott Park Road, CHAD 0377/AP6D-2, Abbott Park, IL 60064-6050 (US).

(72) Inventors: **STEWART, Kent, D.**; 4715 Kings Way North, Gurnee, IL 60031 (US). **STEFFY, Kevin, R.**; 9050 Camino Santa Fe, San Diego, CA 92121 (US). **KEMPF, Dale, J.**; 256 Tyler Court, LIbertyville, IL 60048 (US). **HARRIS, Kevin, S.**; 920 N. Sterling Avenue, Apt. 214, Palantine, IL 60067 (US). **HUTH, Jeffrey, R.**; 1103 Tracy Lane, Libertyville, IL 60048 (US). **STOLL, Vincent, S.**; 218 E. Winchester Road, Libertyville, IL 60048 (US). **HARLAN, John, E.**; 689 Portree Lane, Lake Zurich, IL 60047 (US). **NG, Iok, C.**; 3229 N. Walker E., Arlington Heights, IL 60004 (US). **BETZ, Stephen, F.**; 1136 Lamb Lane, Gurnee, IL 60031 (US).

(74) Agents: **CASUTO, Dianne** et al.; Abbott Laboratories, 100 Abbott Park Road, CHAD 0377/AP6D-2, Abbott Park, IL 60064-6050 (US).

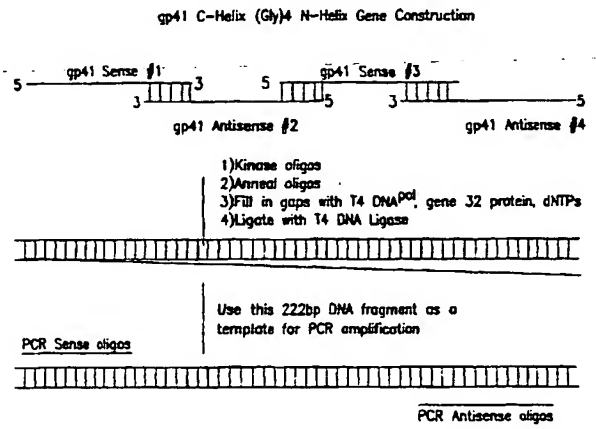
(81) Designated States *(national)*: CA, JP, MX.

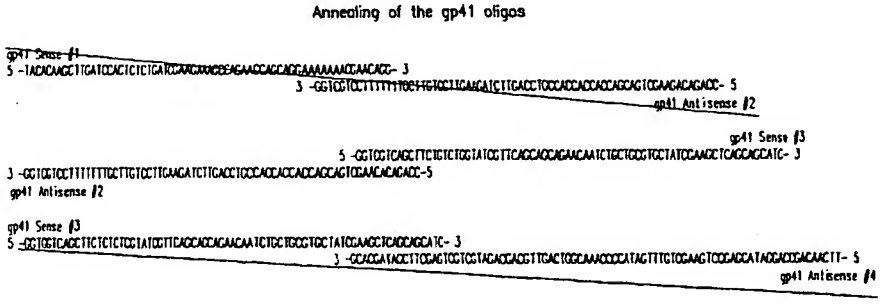
(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*

[Continued on next page]

(54) Title: ENGINEERED CHIMERA OF HIV PROTEIN FRAGMENTS AND USES THEREOF





(57) Abstract: The subject invention encompasses novel proteins related to the human immunodeficiency virus (HIV-1) gp41 protein and to methods of use thereof. For example, the proteins may be utilized in the screening of anti-HIV compounds.

WO 02/034909 A3

with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(88) Date of publication of the international search report:
13 November 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/48040

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/11 C07K14/16 G01N33/569

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, CHEM ABS Data, MEDLINE, SEQUENCE SEARCH, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ECKERT D M ET AL: "INHIBITING HIV-1 ENTRY: DISCOVERY OF D-PEPTIDE INHIBITORS THAT TARGET THE GP41 COILED-COIL POCKET" CELL, CELL PRESS, CAMBRIDGE, NA, US, vol. 99, no. 1, 1 November 1999 (1999-11-01), pages 103-115, XP000982264 ISSN: 0092-8674 cited in the application the whole document --- -/--	1-7,9,10

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 February 2003	05/03/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Niemann, F

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/48040

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	JI HONG ET AL: "Inhibition of human immunodeficiency virus type 1 infectivity by the gp41 core: Role of a conserved hydrophobic cavity in membrane fusion." JOURNAL OF VIROLOGY, vol. 73, no. 10, October 1999 (1999-10), pages 8578-8586, XP002231800 ISSN: 0022-538X the whole document	1-7,9,10
Y	TAN KEMIN ET AL: "Atomic structure of a thermostable subdomain of HIV-1 gp41" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 94, no. 23, 11 November 1997 (1997-11-11), pages 12303-12308, XP002170631 ISSN: 0027-8424 the whole document	1-7,9,10
X	JIANG SHIBO ET AL: "A conformation-specific monoclonal antibody reacting with fusion-active gp41 from the human immunodeficiency virus type 1 envelope glycoprotein" JOURNAL OF VIROLOGY, THE AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 72, no. 12, December 1998 (1998-12), pages 10213-10217, XP002170629 ISSN: 0022-538X abstract	8
A	WO 00 40616 A (WEISS CAROL D ;WILD CARL T (US)) 13 July 2000 (2000-07-13) example 5	

INTERNATIONAL SEARCH REPORT

formation on patent family members

International Application No
PCT/US 01/48040

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0040616 A	13-07-2000	AU 2604200 A	24-07-2000
		CA 2359892 A1	13-07-2000
		EP 1149115 A1	31-10-2001
		WO 0040616 A1	13-07-2000